

SUGGESTED CRITERIA FOR THE EVALUATION OF PHYSICIANS IN FAMILY PRACTICE

DAVID KAPLAN, M.D.

Associate Professor of Medicine
Assistant Dean

MAX WEINER, Ph.D.

Professorial Lecturer in Family Practice

CHARLES M. PLOTZ, M.D.

Chairman, Department of Family Practice
State University of New York Downstate Medical Center
Brooklyn, N.Y.

ONE of the most important yet neglected areas related to the delivery of health care is the evaluation of the quality of care provided by family practitioners. The quality of care is obviously related to the quality of medical training, which in turn depends upon the quality of the faculty in the educational institution. There are two separate concerns: the evaluation of medical practice and the selection and training of faculty. The first task undertaken by the new Department of Family Practice at the medical school was to begin the development of criteria for the selection and training of practicing physicians to serve as preceptors with adjunct faculty rank.

The following criteria were identified and pilot tested for their validity:

- 1) Quality of history taking and physical examination
- 2) Appropriateness of the laboratory and x-ray studies ordered
- 3) Probable correctness of the diagnosis
- 4) Treatment prescribed
- 5) Follow-up schedule recommended
- 6) Appropriateness of consultant use

In addition to the use of specially constructed forms, which objectify the collection of data relating to the criteria, judgments were included

This study was supported in part by a grant from the New York Metropolitan Regional Medical Program, New York, N. Y.

TABLE I. A TYPICAL EVALUATION OF ONE PHYSICIAN PARTICIPATING IN STUDY*

Patient Age	Sex	Hx	Px	Lab x ray	Dx	Rx	Ref	F-Up	O	C	R	S	Diagnosis	Comment
75	F	3	4	3	5	3	5†	5	1	2	2	2	URI	Inadeq. Hx Rx: erythromycin
61	F	5	5	3	5	5	5†	5	3	3	3	2	Angina pect.	
51	M	5	5	5	5	5	5†	5	3	3	3	2	Diabetes mell.	
75	M	5	5	5	5	5	5†	5	3	3	3	2	Diabetes mell.	
36	F	5	5	5†	5	3	5†	5	3	1	1	1	URI: obesity	Amphetamine barbiturate
57	F	5	5	5†	5	5	5†	5	3	3	3	2	Rheum, arth. p. myocard. inf.	
56	M	4	5	5†	4	4	5†	5	2	3	3	2	Angina pect.	
Average		4.6	4.9	4.4	4.9	4.3	5.0	5.0	2.6	2.6		1.9		

*Dr. X....., M.D.; university A.....; 19?? Records: fair; minutes/patient: 6; EFF: 23.

Abbreviations and scoring: The first seven columns refer, in turn, to the quality and pertinence of the history (Hx) and physical examination (Px), the use of laboratory and x-ray studies (Lab x ray), the probable correctness of the diagnosis (Dx), the appropriateness of the treatment (Rx), the appropriateness of whether a consultant was sought (Ref), and the appropriateness of the follow-up schedule (F-Up). Daggers indicate that no studies were ordered and no referral was made and that this was judged appropriate. These parameters were rated on a scale from 1 (very inadequate or inappropriate) to 5 (most adequate or appropriate). Columns headed O, C, R, and S refer to errors of omission (O), errors of commission (C), the requirement for physician's services (R), and the degree of medical sophistication required to manage the patient's problem (S). The Effectiveness Rating (EFF) is the product of the mean scores for errors of omission plus commission (O + C) multiplied by the sum of the mean scores for the requirement for a physician plus degree of medical sophistication required. In this case, for example, $(2.6 + 2.6) \times (2.6 + 1.9) = 23$. URI = upper respiratory infection.

about the degree of medical sophistication required to deal with each patient's problem and whether the patient actually required a physician's services. It was assumed that the data so collected could be used to measure the physician's over-all effectiveness, hence his qualifications as a potential faculty member.

The criteria and forms were pilot tested on a sample of physicians who volunteered to participate in the study. The testing of the forms was undertaken by one faculty member in the Department of Family Practice.

SUBJECTS AND METHODS

In the fall of 1971 New York City physicians listed as family practitioners were invited to apply for participation in the new Department of Family Practice. The respondents, largely from Brooklyn, where the school is situated, were invited for an interview. On the basis of this interview and their *curricula vitae*, seven were selected to participate in a family-practice clinic. All seven volunteered to participate in the evaluation described below. Another 13 were assigned to a second group of future faculty participants. These 13 also agreed to be evaluated in the same way except for one who said his office was "too small." A third group of seven physicians was randomly selected from the original list, all but two of whom were similarly evaluated: one who objected to the concept of peer review and one who said he was "too busy." A total of 24 physicians (22 M.D.s and two D.O.s) were thus evaluated.

THE EVALUATION FORM

It was decided to construct a form which could be used easily by the evaluator. The evaluation form was also constructed so that each physician could be scored for each patient on a scale ranging from one (very inadequate or very inappropriate) to five (very adequate or appropriate). The kinds of data recorded are listed in Table I. From this information a second set of ratings was used to record certain evaluations in consistent summary form (Table I) by means of the scoring indicated below.

Errors of Omission

3 — none

2 — minor (e.g., lack of interest in rectal bleeding which was probably only hemorrhoidal)

- 1 — moderate (e.g., failure to check tendon reflexes in a patient with recent onset of low back pain and sciatic radiation)
- 0 — major (e.g., failure to do urinalysis in a new patient with hypertension)

Errors of Commission

- 3 — none
- 2 — minor (e.g., erythromycin for resolving upper respiratory infection)
- 1 — moderate (e.g., amphetamine + barbiturate for weight reduction)
- 0 — major (e.g., manipulation of metal fragment in cornea)

Requirement for Attention of Physician

- 3 — yes
- 2 — questionable (e.g., could probably have been handled by school nurse; too frequent follow-up)
- 1 — no (e.g., upper respiratory infection for one day in otherwise healthy person; dandruff)

Degree of Medical Sophistication Required

- 3 — expert (a problem which should be handled by a specialist)
- 2 — average (the physician is doing what he is trained to do)
- 1 — minimal (ideally, the task should be performed by a paramedical or non-medical person)

Visits by the faculty member to the physician's office were arranged in advance by telephone. The physician was told the following: that the medical school was setting up a Department of Family Practice, and that fourth-year students participating in the program wished to have the opportunity of spending a month with a family physician; that in order to recruit such physicians, it was necessary that a member of the school visit the office to observe the doctor at work. All such visits were made by one of the physician authors who was constantly with the physician except during occasional pelvic examinations, and in the case of one patient who refused to have the observer present. In addition, in order to obtain some insight into the reliability and reproducibility of the evaluation criteria, three doctors were revisited by a second physician observer.

RESULTS

Table I is an example of the summary sheet compiled on one of the

physicians from data gathered at the time of the office visit. It is self-explanatory and, in general, as will be noted from the data presented below, indicates the performance of an approximately average physician.

History-taking. Of all the parameters measured, the physician's skill at history-taking was perhaps the easiest to evaluate. On a scale of 1-5, the mean score of the 24 physicians was 4.3 (range 1.6 to 5.0). This suggests a rather good level of performance, although often little ability in this area was required. For example, the physician will almost automatically receive a good score when interviewing a patient with an upper respiratory infection or a sprained ankle. Only four physicians earned mean scores of less than 3.5, which might be considered "passing." In only one case did the poor score appear to be related to an inability to obtain a coherent and relevant history. In the others, the poor performance represented a disinclination on the part of the physician to obtain *any* history, apparently because of the time it would have required. Considered in regard to each of the 215 individual patients for whose histories a score could be assigned, the number whose histories were judged inadequate, a score of 1 or 2 on a scale of 5 amounted to 36 patients (17%). Ten patients had histories which, as taken, were judged of equivocal adequacy—a score of 3.

Physical examination. This was difficult to evaluate since, with rare exceptions, the observer did not examine the patient himself. The mean score was 4.0 (range 1.9 to 5.0). Considering the large amount of minimal disease seen, this does not truly represent a good level of performance. That is, among the 207 patients seen by physicians where reasonable judgment could be made regarding the quality of the physical examination, 33 (16%) were judged to be inadequate (scoring 1 or 2 on a scale of 1-5), and 28 others (13%) of equivocal adequacy (scoring 3). The physical examination was most likely to be inadequate in patients with significant illness. While there were occasional instances in which a physician appeared to be trying to perform an appropriate examination and was unable to, more often a poor performance appeared to be related to an unwillingness to spend the time. A striking behavior pattern for several physicians was a kind of ritualistic, semiconscious, brief physical examination performed on every patient regardless of the complaint or the illness. For example, physicians who had expensive wall-attached otoscopes tended to look at the tympanic membranes of all their patients, sometimes not making more relevant observations.

Use of laboratory and x-ray studies. These were generally intelligently and appropriately used by the physicians. Failure to obtain appropriate studies almost never occurred. Poor performance in this area was almost exclusively related to two factors: 1) the acceptance of poor quality x rays taken in the physician's office and 2) blood testing and taking of electrocardiograms rather freely when to the economic advantage of the physician. The mean score was 4.6 (range 3.8 to 5.0). The physicians were judged to have performed inadequately in this area in only eight of 218 instances (4%) and with equivocal adequacy in 23 instances (11%).

Diagnosis. Here again the performance was quite good, although there were only rare diagnostic problems. The mean score was 4.4 (range 3.2 to 5.0). Patients tended to have either minimal disease or, when more than minimal, long-standing, chronic illnesses in which diagnosis was not often a problem on any particular office visit. When diagnostic errors were made they occurred most often in one of two situations: 1) an unwillingness to make a serious diagnosis such as congestive heart failure rather than "swollen ankles" and "age" in order to avoid the problem of administering digitalis rather than treating with a diuretic and vitamin B₁₂ and 2) an unwillingness to take seriously potentially meaningful symptoms or signs, apparently hoping they would go away. Some physicians, like some patients, have strong denial mechanisms.

Treatment. The mean score was 4.2 (range 2.4 to 5.0). The errors made in managing patients were rarely considered serious. (see below, *Errors of commission*). The minor errors were usually made in over-treating patients with minimal or no disease (e.g., antibiotics for viral infections, many months of diathermy for a backache). More serious errors were apt to be made in under-treating more significant disease (e.g., inadequate treatment of hypertension).

Referrals. Referrals to other physicians were made infrequently, and this was usually appropriate, the mean score being 4.8 (range 3.8 to 4.0). It seems somewhat paradoxical that in view of the insecurity some physicians seemed to have in handling certain types of problems, the mistakes were overwhelmingly in the direction of under- rather than over-referral.

Follow-up schedule. The mean score was 4.4 (range 1.5 to 5.0). There were no instances where appropriate follow-up was not suggested. All instances in which the physician's judgment could be ques-

tioned related to patients who probably should not have been seeing a physician at all (follow-up visits for upper respiratory infection or for another injection of vitamin B₁₂) or to patients who were being seen much too frequently at the physician's, not the patient's, request.

Errors of omission. As noted previously, scored on a basis of 3 (no error), 2 (minor error), 1 (moderate error), or 0 (major error), the mean score for the physicians was 2.5 (range 1.7 to 3.0). For 219 patients, a judgment could be made in this regard.

In 136 instances (62%) no error was made.

In 51 instances (23%) a minor error was made (examples: putting off the removal of sutures, somewhat inadequate neurological examination in a patient complaining of dizziness, inadequate history when it probably did not matter very much).

In 27 instances (12%) a moderate error was made (examples: poor history or physical examination when it probably made a difference to proper management, failure to treat or advise about significant hypertension).

In five instances (2%) a major error was made: 1) failure to do urinalysis in a new patient with hypertension, 2) poor history-taking, physical examination, and management of a patient with abdominal pain and bloody diarrhea, 3) totally inadequate management of a patient with congestive heart failure with premature ventricular contractions and a recent myocardial infarction, 4) totally inadequate work-ups of a new patient 87 years old, 5) failure to consider the diagnosis of gonorrhea in a new patient with probable recurrent infection.

Errors of commission. Scored in the same way as were errors of omission, the mean score was 2.7 (range 2.0 to 3.0). The slightly higher score compared to the score for errors of omission, as noted previously, largely reflects the fact that over-treatment of minor illness is not usually as potentially detrimental to the patient as under-treatment of serious illness. For 219 patients, the distribution of errors in this category was as follows:

In 138 instances (63%) none were made.

In 59 instances (27%) a minor error was made (examples: antibiotic for an upper respiratory infection, too many x rays for an old back injury).

In 19 instances (9%) a moderate error was made (examples: amphetamine + barbituate for obesity, insistence, in opposition to the

patient's wishes, of weekly vitamin B₁₂ injections for no illness).

In three instances (1%) a serious error was made: 1) ill-advised attempt to remove metal fragment from cornea, 2) large number of drugs and bacterial antigen injections for minimal obesity and history of respiratory infection (2 cases).

Requirement for a physician's services. As described under Subjects and Methods, on a scale of 3 (required), 2 (questionable requirement), and 1 (not required), the mean score was 2.5 (range 1.7 to 2.8). In 219 instances it could be determined whether the patient appeared to require a physician's services at the particular time that he was seen. The results are as follows:

Yes: 122/219 (56%).

Questionable: 76/219 (35%). Examples: post-ankle sprain, moderate obesity, upper respiratory infection.

No: 21/219 (10%). Examples: band-aid for tiny cut, runny nose for a day, inability to discern any reason for visit.

Medical sophistication required. In 219 instances, a judgment could be made as to the degree of medical sophistication required to manage the patient. The results are as follows:

High degree of sophistication required: 6/219 (3%). Examples: severe Laennec's cirrhosis, rheumatic heart disease with congestive failure and history of emboli, multiple premature ventricular contractions refractory to all usual drugs.

Average degree of medical sophistication required: 161/219 (74%).

Minimal sophistication required: 49/219 (22%). Examples: school check-up, changing bandages, immunizations.

Characteristics of patients. Fifty-seven percent of the patients seen by the physicians were women; 43% were men. The age distribution was as follows:

0-17 years	17%
18-35 years	21%
36-50 years	19%
51-65 years	24%
66+	19%

The major problems of the patients are listed in Table II. In all categories the problems did not generally appear to be particularly serious. Under "heart disease, hypertension, and diabetes," for example, very few patients were taking digitalis or antiarrhythmic drugs, hyperten-

TABLE II. MAJOR DIAGNOSES FOR THE 219 PATIENTS SEEN WHO COULD BE CLASSIFIED*

<i>Adult</i>		<i>Pediatric</i>	
Heart disease, hypertension, diabetes mellitus	62 (34%)	Upper respiratory infection	13 (33%)
Musculoskeletal	24 (13%)	Check-up	10 (26%)
Upper respiratory infection	17 (9%)	Allergy	8 (21%)
Gastrointestinal disease	13 (7%)	Infectious disease	5 (13%)
Check-up	11 (6%)	Immunizations	3 (8%)
Malignant disease	9 (5%)		
Infectious disease	8 (4%)		
Postoperative procedure	7 (4%)		
Psychiatric disease (including two alcoholics)	7 (4%)		
Allergy, asthma	6 (3%)		
Obesity	4 (2%)		
Pregnancy	4 (2%)		
Other	11 (6%)		
Total	183	Total	39

*Three patients had two major diagnoses.

sion was usually benign and treated (often inadequately) with a combination drug, and most diabetics did not appear to require insulin. The musculoskeletal category consisted largely of patients with minor aches and pains or patients who had sustained some trauma in the past and were in the office for diathermy. Gastrointestinal disease was often called "dyspepsia," although there were two patients with melena. The patients with malignant disease, with two exceptions, had had operations and were doing well. No serious infectious diseases were seen except for one patient seen for the first time with a septic abortion and referred back to her original hospital. "Post-surgical procedures" includes changing bandages, putting on band-aids, etc. The psychiatric disease seen was usually managed with vitamin B₁₂, indifference, or hostility; only one psychiatric referral was made.

Over-all effectiveness of the physician. An attempt was made to quantify the general effectiveness of each physician. This parameter would appear to be related to the physician's skills and to the degree that he was required to use those skills. The formula arbitrarily used

TABLE III. SUMMARY DATA OF THE 24 PHYSICIANS EVALUATED*

MD	School	Records	Hx	Px	Lab x ray	Dx	Rx	Ref	F-Up	O	C	R	S	Eff.
1	1	1	4.4	4.0	4.7	4.9	4.6	4.6	4.4	2.7	2.7	1.7	1.3	16
2	2	1	3.8	4.0	4.8	3.4	3.8	3.8	5.0	1.8	3.0	2.8	1.8	23
3	2	2	4.6	4.2	4.1	4.5	4.3	4.8	4.7	2.3	2.8	2.7	1.8	22
4	1	2	4.2	4.0	4.2	4.4	4.8	4.8	5.0	2.2	3.0	2.2	2.0	22
5	1	2	5.0	3.8	5.0	4.8	3.8	5.0	3.5	2.8	2.3	2.5	1.5	20
6	2	2	4.9	4.8	4.8	4.7	4.3	4.7	5.0	2.3	3.0	2.7	1.7	23
7	2	3	4.6	4.9	4.4	4.9	4.3	5.0	5.0	2.6	2.6	2.6	1.9	23
8	2	3	5.0	5.0	4.6	4.6	5.0	5.0	5.0	2.7	3.0	2.7	2.0	27
9	1	3	5.0	3.6	5.0	5.0	4.8	4.8	4.2	2.7	2.7	2.2	1.2	18
10	2	3	4.8	4.9	5.0	4.8	4.9	5.0	4.9	3.0	2.6	2.6	2.0	26
11	1	2	2.5	3.5	4.6	4.4	3.5	4.8	3.6	2.4	2.1	1.9	1.6	16
12	1	3	1.6	2.4	4.3	3.2	2.4	4.3	1.5	1.8	2.0	2.1	2.0	16
13	2	2	4.3	3.1	4.5	4.6	4.3	4.2	4.8	2.0	2.3	2.7	2.0	20
14	2	1	2.1	1.9	3.8	4.2	3.4	4.6	3.5	1.7	2.3	2.4	1.9	17
15	1	2	4.9	4.8	4.8	4.7	3.7	5.0	4.2	2.9	2.3	2.1	1.6	19
16	2	2	5.0	5.0	4.9	4.9	4.6	5.0	5.0	3.0	2.6	2.6	1.6	24
17	2	2	5.0	4.8	5.0	4.7	4.6	5.0	4.8	2.9	2.8	2.4	1.8	24
18	2	3	5.0	5.0	4.9	4.6	4.4	5.0	5.0	2.9	2.5	2.5	2.0	24
19	2	3	5.0	4.7	4.8	4.9	4.9	5.0	4.9	2.9	3.0	2.8	2.1	29
20	2	2	3.4	2.7	4.8	4.9	4.1	5.0	4.4	2.0	2.3	2.5	2.0	19
21	2	3	4.8	4.9	4.8	4.9	4.1	5.0	5.0	3.0	2.5	2.6	1.9	20
22	2	3	4.5	3.6	4.2	4.3	3.9	4.6	4.1	2.4	2.5	2.5	1.9	22
23	2	1	4.3	3.5	4.4	4.0	3.0	4.9	3.6	1.9	2.2	2.7	1.9	19
24	2	3	4.9	4.8	5.0	5.0	4.7	5.0	4.3	2.9	2.5	2.5	2.0	24
Average		2.3	4.3	4.0	4.6	4.4	4.2	4.8	4.4	2.5	2.7	2.5	1.8	21

*See Table I for explanation
Physicians A, B, and C in the text are actually physicians 19, 22, and 23.

for this calculation was $(O + C) \times (R + S)$ (see Table I). Thus, for physician No. 7 in Table I, his effectiveness rating would be $(2.6 + 2.6) \times (2.6 + 1.9) = 23$.

The 24 physicians evaluated in this way segregated themselves into three clusters: Group A (11 physicians) with scores of 16-20, who appeared to be operating at less than desirable effectiveness; Group B (10 physicians) with scores of 22 to 24 who appeared to be generally effective as family physicians, and Group C (three physicians) with scores of 26 to 29 who consistently were seeing sick people and managing them in a superior manner.

Relation of schooling to performance. The physicians could be divided into two groups: Group 1 consisted of 17 physicians who were graduates of American medical schools; Group 2 consisted of seven physicians who had been graduated from foreign medical schools or from an osteopathic school. There was no real difference in the proficiency of the two groups. For instance, the mean score for the sum of errors of omission plus errors of commission for the American graduates was 5.0 (range 4.1 to 5.9), and for the others was 4.9 (range 3.8 to 5.4). As another example, in regard to treatment, the American graduates had a mean score of 4.3 (range 3.0 to 5.0) and the others a score of 4.1 (range 2.4 to 4.8). However, it appeared that the foreign medical graduates and the osteopaths were seeing patients much less in need of physician's services and requiring much less therapeutic skill. For the sum of the scores for $(R + S)$ (see Table I), there was almost no overlap between the two groups—the American graduates having a mean score of 4.5 (range 4.2 to 4.9), the others a mean score of 3.7 (range 3.0 to 4.2). As a consequence, the over-all effectiveness rating of the osteopaths and foreign graduates was significantly less than that of the American graduates, being 18 (range 16 to 22) versus 23 (range 17 to 29) ($p = .01$).

Relation of record keeping with performance. The dogma of medical educators is that good record keeping is the *sine qua non* of good medical practice. An extremely liberal grading system was required in order to divide the physicians' records into three categories: 1) Poor, 2) Fair, and 3) Good. In all parameters there was a relation between ability to keep good records and performance as a physician, but in no cases was there a significant correlation. For example, for records of categories 1, 2, and 3 the sum of the scores of omission and com-

mission were 4.6, 5.0, and 5.2 respectively. Or, as another example, for records scored 1, 2, and, 3, the scores for treatment were 3.7, 4.2, and 4.3 respectively.

Relation of competence ($O + C$) to skill required ($R + S$). One would not necessarily expect a high degree of correlation between these two parameters because some physicians did well with patients with only a questionable need for medical services and requiring little medical skills, whereas other physicians did not do so well with patients who had significant medical problems. There was, however, a significant tendency for physicians to "seek their own level" or for patients to seek a physician appropriate for their needs. Thus, of the 93 patients in whom no errors were made, 60 or 65% required a physician with at least an average amount of skill. Of the 126 patients on whom at least a minor error was made, only 60, or 48%, required a physician with at least average skill ($P < .02$).

Comparison of two observers' evaluations of the same physicians. Three physicians (designated A, B, and C) evaluated by the principal observer (DK) were seen on other occasions by a second observer (BS), who was a chief resident in medicine. A comparison of the scores assigned by the two observers showed good agreement, which suggests that the criteria used to evaluate the physicians' performance are reasonably objective and reproducible.

For example, in regard to competence (the sum of errors of omission and commission) BS assigned physicians A, B, and C scores of 5.3, 5.2, and 3.5. From DK, the respective scores were 6.0, 4.6, and 4.5. For treatment, BS rated them 4.9, 3.2, and 2.7; DK's ratings were 5.0, 3.8, and 3.4. Over-all effectiveness scores were 27.8, 21.8, and 17.1 by BS and 30.0, 20.7, and 18.9 by DK.

Thus, while the absolute scores assigned by the two observers to the three physicians differed, in all parameters both agreed that physician A was superior to physician B, who was superior to physician C.

DISCUSSION

The only firm conclusion which may be drawn from this study is that physicians are not reluctant to have another physician present in their offices while they are seeing patients. This was also true for the patients involved. Even though the setting of a physician observing a colleague may be considered artificial, in as little as two hours the

observers felt they were able to judge the quality of health care delivered.

The primary weakness of this study is that it represents the effort of physicians trained in internal medicine to evaluate the performance of family physicians in relation to only a vaguely conceived concept of what these physicians should be doing. There are no accepted standards of the "good" general practitioner. Admittedly, therefore, it does require a certain amount of arrogance for one physician to determine that 11 of 24 other doctors are operating at less than desirable effectiveness.

Much of what the general practitioner does is usually thought to be related to two areas: 1) health maintenance and 2) the management of functional complaints or anxiety. Our data, and that of others* suggests that this is probably true. The question, however, is whether the playing of a role by the physician automatically legitimizes that role. There were frequent instances of physicians encouraging the dependence of patients. There was always, at the very least, implied acceptance of the patient who came to have his runny nose evaluated, his bandage changed, or his monthly electrocardiogram performed. That such patients usually felt reassured after such visits hardly leads to the conclusion that the physicians are doing what they are trained to do.

On the other hand, it was thought that 13 of the 24 physicians were, in general, practicing at a satisfactory or at a superior level, and were using skill and judgement in the management of significant medical problems.

A word should be said about the homogeneity of scores earned on the various scales. It is the view of the authors that the nature of health-care delivery by the family physician is such that one could expect that a majority of the items evaluated would be rated 4 or 5 on a scale of 5 which, of course, was the case in the present study. Very few physicians in a typical practice are so inadequate that an over-all rating of 3 or lower would be expected to occur. In view of the very high scores, it became readily apparent to the authors that future research which measures health-care delivery should be concerned with individual low-score items as discriminating among physicians. For example, there were a number of subjects who in fact earned 2s or 3s

*Clute, K. F.: *The General Practitioner*. Toronto, Canada, Univ. of Toronto Press, 1963, chap. 15.

but whose over-all score, due to the large number of variables, ended up with a respectable 4 +. It is suggested, therefore, that future evaluators concentrate on those items which result in scores at the lower end of the continuum for a given physician and attempt to identify the quality of health-care delivery on the basis of these items rather than over-all score items.

SUMMARY

This study is a preliminary attempt to quantify the performance of a group of family physicians practicing in a large urban center in order to judge the quality of medical care delivered by these physicians, for use in selecting faculty. A similar study, utilizing a larger group of physicians and patients, with additional observers, might suggest where changes in either the physician's training or his professional activities might be indicated.

Twenty-four family practitioners were observed, while seeing patients in their offices, for the purpose of selecting preceptors for a new Department of Family Practice at a medical school, and for the purpose of evaluating predetermined criteria designed to characterize the quality of care dispensed by these practitioners. It was concluded that 13 of the 24 physicians were, generally, seeing people who required medical services and managing them well, and that 11 physicians were often either seeing people with only a questionable need for a physician's services, or not managing significant illness appropriately. Two relations were noted: 1) The more competent physicians tended to see a higher proportion of patients with significant illness. 2) Graduates of American medical schools tended to function at a more effective level than did graduates of foreign or osteopathic schools.

ACKNOWLEDGMENTS

We deeply appreciate the cooperation of the family physicians involved in this study. They were uniformly gracious and generous. We are grateful also to Dr. Barry Schwartz (BS) for his highly competent assistance.